DOCUMENT RESUME

ED 051 298 TM 001 619

AUTHOR	Womer, Frank B.
TITLE	Measurement in Education: National Assessment Sa[illegible].
INSTITUTION	National Council on Measurement in Education, Ea[illegible] Lansing, Mich.
PUB DATE	Oct 70
NOTE	8p.
AVAILABLE FROM	National Council on Measurement in Education, Office of Evaluation Services, Michigan State University. East Lansing, Michigan 48823. $2.00 per year (4 issues); single issues 0.25 each in quantities of 25 or more
EDRS PRICE	EDRS Price MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS	Adults, *Attitudes, *Changing Attitudes, Citizenship, Data Collection, Evaluation, Evaluation Techniques, *Knowledge Level, Longitudinal Studies, Measurement Techniques, Multiple Choice Tests, *National Surveys, Sciences, *Skills, Standardized Tests, Test Construction, Young Adults, Youth
IDENTIFIERS	Item Difficulty, *National Assessment

ABSTRACT

National Assessment is a data gathering project designed to provide information, in ten subject areas, about knowledge, skills, understandings and attitudes of young people in this country, and to assess changes in these variables over time. The data is collected and reported at the item level. Each exercise was developed with emphasis on content validity, and is geared to sample a specific objective within a subject area. A striking feature of the first National Assessment report is that there are no scores or norms with which to compare results. Instead the individual exercises with the percent choosing or producing each response (p-values), both correct and incorrect, are given. This technique allows the reader to evaluate results and draw inferences for himself rather than just review an average or summary. Also, by looking at the P-values of wrong responses, considerable light may be shed on commonly held misconceptions. Generalizations discussed are based on exercises from the subject areas of Science and Citizenship, with only partial National Assessment results available for the latter, and are drawn by looking at the exercises as a total set of exercises, not as a total score. They are not to be construed as representing the National Assessment's viewpoint, as the selection of these generalizations, rather than others that might be drawn from the data, is a personal one. (Author/CK)

ED051298

Volume 2 No. 1

NCME

October, 1970

measurement in education

A SERIES OF SPECIAL REPORTS OF THE NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

National Assessment Says

Frank B. Womer

ABOUT THIS REPORT

Few educational projects have endured the controversy and demonstrated the promise of National Assessment. An early fear was that National Assessment would be used by "central authorities" to rank the quality of local schools. The project guards against such misuse by protecting the anonymity of individual schools and systems. But this educational census does make it possible for local educators to make local studies of achievement. This is a major strength of the undertaking.

Local and national judgment of the effectiveness of American education is greatly facilitated by National Assessment's concentration upon individual items and their relationship to school objectives. This means of reporting data puts the emphasis on educational outcome in terms that have some absolute meaning. And it makes it possible to take into account local conditions and local goals.

In this report Dr. Womer cites some preliminary results of National Assessment to illustrate the important contribution it can make to education. As staff director he is a well-informed spokesman. Dr. Womer has had a distinguished career as school teacher, editor, and professor at the University of Michigan. He is well-known for his writings and professional services—especially as past President of NCME.

WWW

TM 000 676

FRANK B. WOMER

More young adults between the ages of 26 and 35 (9 out of 10) are aware of the fact that the President does not have the right to do anything affecting the United States that he wants to do than are 17-year-olds (8 out of 10), 13-year-olds (7 out of 10), or 9-year-olds (5 out of 10). The question on which these results are based and the results themselves are presented in Example I in the box on page 3. It comes from a Citizenship exercise for National Assessment. While most of the young adults in the National Assessment sample could state an acceptable reason for their answer (8 out of 10), the younger age groups did not do as well (only 2 out of 10 of the 9-year-olds). These results suggest that for this specific bit of information there is continuing growth through the school years and even into young adulthood.

National Assessment* is a data gathering project designed to provide specific information about knowledges, skills, understandings, and attitudes of young people in this country. The data are collected and reported at the item level, with each item geared to sample a specific objective within a subject area. This information provided by National Assessment has not been available previously. Example I is one exercise released in the first National Assessment reports which covered Science and part of Citizenship results. Many additional reports will be forthcoming over the years — in these two subject areas as well as in eight others now scheduled.

Perhaps the most striking feature of National Assessment's first reports is that there are no scores or norms — just individual exercises (questions, items) along with the percent choosing or producing each response (p-values), both correct and incorrect, for each exercise. Those who developed the plan for National Assessment felt that the best way to describe what young people know is to

*Womer, Frank B. *What Is National Assessment?* Ann Arbor, Michigan: National Assessment of Educational Progress, 1970.

present the questions or tasks that they were asked along with information about how well they performed. This directs attention at actual samples of behavior rather than at some summation of behaviors. It allows the reader of the reports to make his own evaluation of each exercise. He can accept the results as meaningful information useful in teaching and/or curriculum evaluation and/or policy making and/or allocation of educational funds, etc. He can reject a question as meaningless or inappropriate iff his judgment leads him to that conclusion. The point is that the reader of a report has all of the results before him rather than an average or a summary or a conclusion.

With this type of reporting in mind, National Assessment developed its exercises with an eye to content validity, as judged by subject matter specialists, other educators, and laymen. The exercises were not item analyzed (there is no total score) nor were they related to future performance (there are no criterion measures). The purpose of National Assessment exercises in toto is to describe, by example, what most young people know and can do, what about half can do and what very few can do. The purpose of a single exercise is to stand as one example of a meaningful knowledge or skill or attitude that relates to a specific objective in a given subject area.

This type of report is "dangerous" because it exposes each question or task to critical examination by a reader. Each of the nearly 200 Science exercises released in 1970 is subject to scrutiny for individual imperfections. And even after being reviewed by between 12 and 20 persons, some exercises are still less than perfect. But this type of report also is "courageous" because it allows acceptance or rejection of each exercise individually. A reader is not presented with generalizations based on materials that he has never seen.

Another striking feature of National Assessment's initial reports is that there are no standards or norms against which one can compare the results. Consider Example II. In that Science exercise two-thirds of the 17s and half of the adults knew the correct answer (indicated by the blackened circle). But is two-thirds good or bad? Is half good or bad? There is no statistical reference point. Several science educators who reviewed these results have indicated dissatisfaction that more respondents did not know about the interrelationship of animal and plant life in an ecosystem. But such judgment relates to an expectation, an internal personal standard. And it is exactly this type of standard, personal judgment, that must be used to draw certain types of conclusions from the results — particularly the type of conclusion that attempts to judge whether or not young people are learning what they "should" learn. Should, in this context, must be a personal, thoughtful judgment.

One of the initial reviewers of Example II was not disappointed because of any feeling that more 17s and more adults should know about ecosystems. But she was disappointed by the results of this exercise because, in her judgment, it is more of a reasoning exercise than a knowledge exercise. And she felt that more 17s and more adults should have been able to determine the answer by logical deduction from the information given. This illustrates the fact that different readers will have different insights into National Assessment results.

Eventually National Assessment will generate its own standards, although not in the same sense as those established for the usual standardized test. Only half or fewer of the exercises administered by

NCME

measurement in education

Volume 2, No. 1 October, 1970

Warren W. Willingham, Editor

Advisory Committee:
J. Thomas Hastings, University of Illinois
Alice J. Irby, Educational Testing Service
Robert S. Lankton, Detroit Public Schools

Measurement in Education is a series of special reports published four times a year in October, January, March, and May by the National Council on Measurement in Education. These reports are concerned with the practical implications of measurement and related research and their application to educational problems of individuals, institutions, and systems. The emphasis is upon uses of measurement rather than technical or theoretical issues. Subscription rate: $2.00 per year; single copies 20¢ each in quantities of 25 or more.

Second class postage paid at East Lansing, Michigan 48823

Address Editorial Correspondence to:

Warren W. Willingham
College Board
800 Welch Road
Palo Alto, Ca. 94304

Address Business Correspondence to:

Irvin J. Lehmann
Office of Evaluation Service
Michigan State University
East Lansing, Mich. 48823

EXAMPLE I

A. Does the President have the right to do anything affecting the United States that he wants to do? (Yes, No, I don't know)

B. (If yes) Why? (Part B was not scored; it was asked to insure that respondents understood Part A and to give them a chance to explain their position.)

C. (If no) Why not?

(If answer to C is vague) Who or what would stop him from doing what he wants?

Acceptable reasons to C (examples): People could stop him; elected officials could stop him; checks and balances system of government; laws stop him; country would be a dictatorship; not the democratic way.

Unacceptable reasons to C (examples): Police or Vice President would stop him; he wouldn't be doing his job; he might do something that could hurt the country; he would be doing what is right; people vote for him not to; he can't do it; everybody, even the President, has some limitations; he just advises us; he can't do everything since he is only one person.

Results

	Age 9	13	17	Adult
Stated that the President does *not* have the right to do anything affecting the United States that he wants (No to A)	49%	73%	78%	89%
Stated that the President does not have the right and gave an acceptable reason (acceptable reason to C as well as No to A)	18	53	68	80

National Assessment in a given year are released that year. The others are retained to be used again three or six years hence. At that time it will be possible to compare the results from a second or third assessment with those obtained previously. Then one can see whether change (progress?) is taking place in the knowledges and skills of young people over time. This is the ultimate goal of National Assessment — to measure changes in knowledges, skills, and attitudes over time.

Since the first results of National Assessment are benchmark data, they provide neither instantaneous "indictments" of American education nor instantaneous "whitewashing." This fact has been of considerable disappointment to persons who looked upon the project as one which will provide "answers" to all sorts of educational questions. An information-gathering project is not designed to provide such answers. But it can and should provide decision-makers with information useful in decision-making. Hopefully National Assessment will do just that.

Even though the ultimate goal of National Assessment is to assess change, the first results do point to some generalizations of considerable import, as well as illustrating specific knowledges and skills and attitudes that young people have and have not attained. The generalizations discussed here are based on total national results for Science and on partial national results for Citizenship. Later reports will include comparisons, item by item, for four geographic regions, size and type of community, sex, color (black versus non-black), and an educational index of the home.

EXAMPLE II

In a particular meadow there are many rabbits that eat the grass. There are also many hawks that eat the rabbits. Last year a disease broke out among the rabbits and a great number of them died. Which of the following probably then occurred?

Results

Age 17	Adult		
4%	2%	○	The grass died and the hawk population decreased.
1	1	○	The grass died and the hawk population increased.
68	52	●	The grass grew taller and the hawk population decreased.
4	4		The grass grew taller and the hawk population increased.
20	30	○	Neither the grass nor the hawks were affected by the death of the rabbits.
2	10	○	I don't know.
1	1		No response
100%	100%		

The following generalizations are based on looking at the exercises as a total set of exercises, but not as a total score. An attempt is made to identify what the data say versus what the author says. The author's views are not to be construed as interpretations representing a National Assessment viewpoint, and the selection of these generalizations rather than others that might be drawn from the data is a personal one.

The evidence for this statement is based on "overlap" exercises, those administered to more than one age level. Consistently on the overlap exercises 17s did better than 13s and 13s did better than 9s. There were 15 overlaps between 9s and 13s for Science and 17 for Citizenship. All of the Science overlaps and 13 of the 17 Citizenship overlaps favor the 13s over 9s. There were 23 overlaps between 13s and 17s for Science and 73 for Citizenship. All of the Science overlaps and 47 of the 73 Citizenship overlaps favor the 17s over 13s. Note that the generalization does not take a position as to whether the schools, or other social organizations, or the family, or any other causal factor or combination of factors are responsible.

A statement that learning is taking place as young people proceed through the school years is not exactly revolutionary. Observation and common sense have indicated as much. But National Assessment documents this generalization in terms of specific knowledges and skills. For example, when students were asked to read a chart depicting seven weights, 80 percent of the 9s were able to identify the greatest weight, whereas 92 percent of the 13s were able to do so. (There was a single number that was greatest – 64 pounds.) From the same chart 54 percent of the 9s could identify the smallest weight (distinguish between 2 pounds and 2 ounces) whereas 81 percent of 13s could do it. When asked about what happens when scientists carefully measure any quantity many times, 69 percent of the 13s correctly selected the alternative "most of the measurements will be close but not exactly the same" while 72 percent of the 17s did so. When asked to write the names of the president, vice-president, secretary of state, secretary of defense and five other federal office holders, 17s did better than 13s for each of the nine positions (ranging from 2 percent of the 13s who could write McCormack as the speaker of the house to 98 percent of the 17s who could write Nixon as the president).

Perhaps the greatest utility of this specific type of information from overlap exercises will be as an aid to understand when growth is taking place in specific skill and knowledge areas. The difference of 27 percentage points between 54 percent and 81 percent for the one exercise quoted above is substantial. The difference of three percentage points between 69 percent and 72 percent is not. If one wanted to generalize from this to a statement about the types of scientific knowledges on which 13s do better than 9s or 17s better than 13s, a much larger number of exercises would be needed.

It may be noted that all of the reversals in the exercises initially reported are in Citizenship. For four of the overlaps 9s did better than 13s and for 26 overlaps 13s did better than 17s. One also may note that many of the Citizenship exercises are attitudinal whereas most of the Science exercises require knowledges and skills. But again, until all of the Citizenship exercises are analyzed it would be hasty to come to any conclusion on this matter.

There were 58 Science and 57 Citizenship exercises administered both to 17s and adults. Of these, 38 Science and 10 Citizenship overlaps yielded higher p-values for the correct answer for 17s, while 20 Science and 47 Citizenship overlaps yielded higher p-values for adults. Because the 17s did better on more Science exercises than the adults one might be tempted to assume general superiority of the 17s. However, that would be a hasty and unsupported generalization. Careful examination of the overlap exercises has led several reviewers to the conclusion that the exercises for which 17s did better tend to be of a different type than the ones for which adults did better. The Science report states: *

> 'Examination of the released exercises suggests that Adults do as well or better than 17s when asked questions which they may know from personal experience, whereas 17s do better on exercises which require formal education. Thus, Adults do better than 17s on two exercises which call for knowledge on human reproduction and on an exercise about fuses. On the other hand, more 17s than Adults successfully chose the response 'electrons' when asked, 'An electric current in a copper wire involves mainly the movement of . . .' and given five alternatives. Similarly, 17s were more successful on: 'Two light waves are traveling in a vacuum. The wave with the higher frequency will have the (shorter wave length).' While 32 percent of the 17s chose the correct response, 20 percent of the Adults chose it."

*_Science: National Results, 1970._ National Assessment of Educational Progress: Ann Arbor, Michigan, 1970.

"The seven released exercises that were answered correctly by most Adults deal with non-technical information that might be found in newspaper or magazine articles dealing with scientific topics or in television programs on science. Five of the seven might be considered to have to do with biology or medicine.

"The exercises that few Adults answered correctly are quite technical in nature, involving knowledge that is likely to be learned only in school and is reinforced by experience by few young adults (e.g., the periodic table)."

From these remarks it must be concluded that a different balance of Science exercises could have yielded a different balance of results for 17s and adults. Although adults answered correctly more often than 17s in most of the Citizenship overlaps, the exercises in the first partial results are heavily attitudinal or deal with knowledge of government. When all of the Citizenship results are available this trend may or may not be evident.

The generalization that young adults sometimes show greater achievement than 17s in Science and Citizenship may, like the first generalization discussed, seem to many readers to be another bit of "common sense", but documentation of common sense can have its own utility. Documentation can focus attention upon an area of learning in a way that common sense may not. This simple generalization helps to remind us that much learning takes place outside of schools, that "textbook" learning may have limited utility (if textbook implies rote memory primarily), and that if the ultimate goal of education is an enlightened citizenry one needs to examine carefully what knowledges and skills adults truly need to acquire *and retain.*

III. There are meaningful knowledges and skills and attitudes in Science and Citizenship at all difficulty levels.

In producing the exercises to be used in National Assessment, item writers were given several criteria to follow. Among them were: 1) to produce exercises with high content validity which relate to specific objectives, and 2) to produce exercises that were very easy (one-third), very difficult (one-third), and in-between (one-third). Thus, Science item writers were asked to produce exercises that almost all young people could answer correctly and that were criterion-referenced and meaningful, to produce very difficult exercises that also were meaningful, and so on. The primary problem here was to sample each difficulty level without artificial manipulation. When pressed by the author to produce a greater number of easy exercises, one writer responded by saying that he could do it easily by substituting ridiculous foils for logical foils in multiple-choice exercises. This response ignored the criterion of high content validity.

If item writers had been able to do exactly what had been asked of them, the results would have produced trimodal distributions of p-values. What were in fact produced were essentially rectangular distributions, definitely *not* normal distributions of p-values (see pages 8-11 of *Science: National Results, 1970*). For ages 9 and 13 there was a deficiency of very difficult exercises ($p<.25$) whereas 17 and adult results were better spread over the total range of p-values. This may seem surprising when one considers that most of the Science exercises were multiple-choice. One might have expected truncated distributions with very few exercises with p-values below the chance probability levels, but this did not occur. The question of chance was discussed at great length prior to the first assessment. In an attempt to minimize guessing (and based on specific research results) an "I don't know" choice was added to almost every multiple-choice exercise. It was elected as an option by quite a few respondents, particularly for the very difficult questions.

But what is the import of the generalization that there are meaningful knowledges and skills at all difficulty levels? It suggests that all young people have acquired meaningful knowledges and skills that relate directly to objectives of instruction in Science and in Citizenship. Most people probably would have paid lip service to this statement prior to any results of National Assessment, but unfortunately too many of us (including we teachers, who should know better) have acted as if we felt that some youngsters were completely devoid of useful skills or knowledges. As National Assessment results are accumulated over the years, it should be possible to develop a picture of what knowledges and skills all 9s, 13s, and 17s have attained. Whether we will be satisfied with that picture is another question, but at least we will know where students stand. This should be of considerable help in planning for group learning experiences, in avoiding knowledge already acquired and in building knowledge not yet acquired.

The results in Example III suggest that society has done a fairly good job in getting young people to indicate lack of bias toward people of other races, *in a paper and pencil situation.* The obvious response to this is that what people say they would be willing to do and what they really do may not be the same. Nevertheless, if young people did not indicate tolerance, there would be little chance for further progress in race relations. To the author the most disturbing aspect of Example III is that

EXAMPLE III*

People feel differently toward people of other races. How willing would you be to have a person of a different race doing these things?

Results

[For each situation below, the choices were: Willing to, prefer not to]	% willing to Age 13	Age 17	Adult
A. Be your dentist or doctor?	81%	74%	75%
B. Live next door to you?	83	77	67
C. Represent you in some elected office?	81	82	82
D. Sit at a table next to yours in a crowded restaurant?	80	90	88
E. Stay in the same hotel or motel as you?	88	92	89
Willing to for one or more of the above			
two or more . . .	96	97	93
three or more . . .	94	94	90
four or more . . .	89	88	86
all five . . .	56	56	56

*Not administered to the in-school sample in one large western state, one southeastern county and one south-western city at the request of state or local authorities.

almost half (44 percent) of the respondents at each age level did indicate an unwillingness to accept a person of another race on at least one of the five categories.

Samples of information known to very few young people present a situation of a different sort. They must be related to some standard of whether we would (should) ever expect large numbers of students to acquire that skill or knowledge. For example, one might not be at all concerned that only 6 percent of 17s can identify tin and sulfur as the two elements which have been oxidized, when shown a specific chemical equation. On the other hand one might be quite disturbed (as several reviewers were) that only 7 percent of 9s could correctly answer the Science Exercise in Example IV. In this Exercise most 9s simply added

the two temperatures rather than taking the average. Whether this is because 9s don't understand the concept of an average, whether it is because word problems bother them, whether the word Fahrenheit was a problem for 9s, or whether there is some other reason, I suspect most of us would feel that this is the sort of skill that must be acquired by more than 7 percent of our young people, at some appropriate age.

This generalization about the breadth and depth of knowledges that young people have will hearten some and frighten others. Such information can help to focus our attention where it ought to be in education — on what young people are or are not learning and on what attitudes they are or are not developing. And this is accomplished with specific examples rather than summary statistics.

When item writers prepared exercises for National Assessment, they were asked to classify each one as 10 percent (very difficult), 50 percent (moderately difficult), or 90 percent (very easy). As various subject matter reviewers examined the exercises, they were asked to note instances where they disagreed with the writer's estimate of difficulty. Some of the 10-50-90 designations were changed; most were not.

EXAMPLE IV

A pint of water at a temperature of 50° Fahrenheit is mixed with a pint of water at 70° Fahrenheit. The temperature of the water just after mixing will be *about*

Results

Age 9		
4%	○	20°F.
2	○	50°F.
7	●	60°F.
5	○	70°F.
69	○	120°F.
12	○	I don't know.
0		No response
99%		

When reporting the results for Science, three categories of correct responses were established: rather few (0-33 percent), good many (34-66 percent), and most (67-100 percent). The original writers' and reviewers' estimates have now been plotted against the actual results. Of 498 comparisons made for Science, 339 (68 percent) were the same whereas 159 (32 percent) were different. Thus the item writers and reviewers judged difficulty correctly two-thirds of the time and judged incorrectly one-third of the time. In the author's opinion, this is not outstanding success. This is not to suggest that the particular writers and reviewers for National Assessment were poor judges, since it is not known whether other writers could have done better or whether writers in other subject areas could have done better.

Across the four age groups the percentages of agreement were 70, 65, 70, and 67 for ages 9, 13, 17 and adult. Thus, judgments were quite similar across ages. The writers were correct 80 percent of the time for the easy exercises (the 90s) whereas they were correct only 60 percent of the time for the others (50s and 10s).

These results raise the question of whether it is possible for adults to do a really good job of estimating item p-values of students. One early critic of National Assessment suggested to the author that since we were asking writers to prepare exercises with difficulty levels of 10, 50, and 90, we really did not need to collect data. The results, however, indicate that in the final analysis it is 9-year-olds who must tell us what 9-year-olds know, that it is 13-year-olds who must tell us what 13-year-olds know, and so on.

Of the 49 overlap exercises between ages 17 and adult which included "I don't know", adults chose the "I don't know" more often than 17s for 41 exercises while 17s used it more often for the other eight. One might assume, since 17s actually did better on more exercises than adults, that the adults were just realistic about their lack of knowledge. The results do not support this assumption, however, since 17s were also wrong more often than adults (determined by adding p-values for incorrect foils). Thus, it is safer to conculde for the 49 overlap Science exercises that adults are less willing to guess than 17s. Why this is so is another question. The author's personal hypothesis is that school-age youngsters are so geared to guessing on multiple-choice exercises (and rightfully so) that many of them never seriously considered the "I don't know" alternative. Adults, who are assessed in their own homes, probably are less concerned about a "score" that they might be achieving than about giving their best response or straightforwardly admitting a lack of knowledge. Whether it is good or bad that 17s seem to be prone to more guessing than adults is a debatable question.

The fact that trained item writers and reviewers were "surprised" with respect to estimated difficulty fully one-third of the time is strong support for this statement. Noneducators probably will be surprised even more. Evidence for this assumption comes from the initial newspaper articles written about National Assessment results. The education writer for The Washington Post was unhappily surprised that, while 70 percent, 91 percent, and 92 percent of the 13s, 17s, and adults respectively knew that the Senate was the second of the two houses of Congress, 17 percent, five percent, and four percent respectively *thought* that the Supreme Court was the answer. From a psychometric viewpoint, any p-value above 90 percent might seem to be good, but to a Washington reporter anything less than 100 percent on such an item is unthinkable. The same reporter was disturbed that only half of the 17s and adults could solve this problem correctly: "A motor boat can travel five miles an hour on a still lake. If this boat travels downstream on a river that is flowing five miles per hour, how long will it take the boat to reach a bridge that is 10 miles downstream?"

One reviewer of the National Assessment Science results finished his comments by listing eight pleasant surprises and 10 unpleasant surprises. He was pleased that 89 percent of 17s knew that living dinosaurs have never been seen by man ("The Flintstones not withstanding") and was displeased that only 33 percent of 17s and 25 percent of adults knew that doubling the linear dimensions of a cube increases its volume eightfold.

Readers of the National Assessment reports may want to play the same game — estimating what they think students know and can do in specific exercises before looking at the p-values. Another potentially fruitful approach for teachers and curriculum specialists is to look at the p-values of wrong responses as well as of the correct responses. Such analyses have the potential of shedding considerable light on specific misconceptions that are commonly held. For teachers, broad generalizations that may be abstracted from National Assessment results may be of much less interest than specific, item-by-item analyses. Such analyses probably are best done by subject matter specialists rather than measurement personnel, though certainly most NCME members could have considerable positive input to such analyses.

7

In order to illustrate future reports, 10 Science exercises for age 17 were broken down along six dimensions (main effects). Sometime in the first half of 1971 all released exercises will be reported this way. Ten exercises alone are too few to use to draw any conclusions. Yet we may note that for seven of these 10 exercises, 17s whose parents had some high school work but did not graduate were significantly (more than 1.5 standard errors) below the national p-values. In contrast, for 8 of the 10 exercises the 17s whose parents had taken work beyond high school graduation scored significantly above the national p-values. For 17s whose parents graduated from high school but had not taken additional work, only one of the 10 exercises showed any significant variation from the national values. It will be interesting to find out whether this trend continues across all of the Science exercises.

In some instances it may prove advantageous to look separately at those exercises that show great differences between groups and those that do not. Such analyses for blacks versus non-blacks may prove useful if they enable one to take a look at where achievement differences seem substantial versus where they are minimal. For example, an exercise requiring actual manipulation of weights on a balance beam was answered correctly by 50 percent of the blacks and 78 percent of the whites. In contrast the following exercise was answered correctly by 55 percent of the blacks and 54 percent of the non-blacks: "A five-pound rock is dropped from a cliff 500 feet high. The longer the rock falls, the greater is its (speed)". Analyses and evaluation of results such as these could keep educators, curriculum specialists, and sociologists busy for some time.

The purpose of this article has been to present some of the highlights of the first National Assessment results. After five years of preparation and a year of data collection, it now is possible to see the complete national results for Science and part of the national results for Citizenship (the rest are due in print this fall or winter). A major step forward in information gathering in education has been taken. If this new information is to prove useful and helpful to laymen and educators, it must be widely disseminated, discussed, analyzed, evaluated, cursed, and praised. NCME members are in a unique position to influence the ultimate success or failure of National Assessment by individual dissemination in schools (workshops, institutes, curriculum libraries, etc.) and colleges (courses, libraries, workshops, etc.). Inquiries regarding National Assessment reports should be addressed to the Superintendent of Documents, Government Printing Office, Washington, D.C. Additional information about the National Assessment project can be secured from 2222 Fuller Road, 201A Huron Towers, Ann Arbor, Michigan, 48105.

*

NCME NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

Second class postage paid at
East Lansing, Michigan 48823

Office of Evaluation Services
Michigan State University
East Lansing, Michigan 48823

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY

NCME

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

A series of special reports of the National Council on Measurement in Education